DOCUMENT RESUME

ED 069 739 TM 002 194

AUTHOR	Medley, Donald M.; Quirk, Thomas J.
TITLE	Race and Subject-Matter Influences on Performance on General Education Items of the National Teacher Examinations.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-72-43
PUB DATE	Sep 72
NOTE	42p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*General Education; *National Competency Tests; Negro Culture; Performance; *Racial Factors; *Standardized Tests; Subculture; *Teacher Evaluation
IDENTIFIERS	*National Teacher Examinations

ABSTRACT

Race and subject-matter influences on performance on the National Teacher Examinations (NTE) were assessed in a study of the effects of changes in content of so-called general education items designed to reflect contributions of black and "modern" subcultures. The test used in the study was a 65-item subtest of the NTE designed to measure candidates' knowledge of subject-matter content in the areas of social studies, literature, and fine arts, sometimes called the general education subtest. Original and replication studies were made and showed consistently the impact on relative performances of black and white candidates that relate to decisions about relative numbers of black, modern, and traditional items included in the NTE. The results leave little doubt that black candidates tend to possess a different set of knowledge than white candidates, and that these differences have little to do with conventional subject-matter areas. (LH)

ED 069739
TM 002 194

RB-72-43

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RESEARCH BULLETIN

RACE AND SUBJECT-MATTER INFLUENCES ON PERFORMANCE ON GENERAL EDUCATION ITEMS OF THE NATIONAL TEACHER EXAMINATIONS

Donald M. Medley
University of Virginia

and

Thomas J. Quirk
Educational Testing Service

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
September 1972

TABLE OF CONTENTS

	Page
Rationale	1
Description of the Experimental Forms	2
PROCEDURE	4
The Sample of Candidates for the Original Study	4
The Sample of Candidates for the Replication Study	6
DATA ANALYSIS PROCEDURES	9
RESULTS	12
The Original Study	12
Replication of Form A	15
Form B	18
SUMMARY AND DISCUSSION	19
REFERENCES	22
FOOTNOTES	23

RACE AND SUBJECT-MATTER INFLUENCES ON PERFORMANCE ON GENERAL EDUCATION ITEMS OF THE NATIONAL TEACHER EXAMINATIONS[1]

The primary purpose of this study was to assess race and subject-matter influences on performance on the National Teacher Examinations. Specifically, the effects of changes in content of so-called general education items designed to reflect contributions of black and "modern" subcultures were investigated. The test used in the study was a 65-item subtest of the National Teacher Examinations (NTE) designed to measure candidates' knowledge of subject-matter content in the areas of social studies, literature, and fine arts, sometimes called the general education subtest.

Rationale

The basic question asked was whether, and to what extent, projected changes in the test would alter the relative standings of candidates in the population that takes the examination. If changes in item content do produce systematic changes in rank order, then the validity of the test (for any purpose) may be affected, since the type of candidate who does best on the examination will change.

Hypotheses about what type of candidate is likely to perform differently on some new kind of item may be tested by studying how the mean relative scores of groups of candidates are affected by changes in item content. Subtest scores on the Social Studies, Literature and Fine Arts section of the NTE have never been equated from form to form, thus it is not possible to generalize correlations between this subtest and "teacher effectiveness," even if a reasonable criterion of teacher effectiveness could be presented. Given these two factors it is not possible to evaluate systematically the possible bias of the test within a predictive framework. The differences between the performance of known subgroups of candidates on items of different types, therefore, become important in their own right.

Description of the Experimental Forms

Two experimental forms of the NTE Common Examinations were prepared (Forms A and B). Form A was administered during one of the regular administrations of the NTE in 1970 and again during one of the regular NTE administrations in 1971, while Form B was used during that same regular administration of the NTE in 1971. The second form was used to study whether the effects detected in 1970 could be replicated with different items as well as with different groups of candidates. The Social Studies, Literature, and Fine Arts section of each form was modified to include a larger proportion of items than the regular form in terms of items related to contemporary culture ("modern" items) and of items related to cultural contributions of black authors and artists ("black" items), and a smaller proportion of items than the regular form in terms of items related to classical and traditional culture ("traditional" items).

Insert Table 1 about here

In each of the two experimental forms, enough experimental items were substituted for traditional items within their corresponding subtests on the regular form of the examination to achieve the distribution of item types shown in Table 1. Neither the experimental items nor the items which they replaced were used in determining a candidate's NTE score, so that the NTE scores of a candidate would be based on the same items regardless of whether he took the regular form or one of the two experimental forms of the test.

The black and modern items written for the experimental forms were not designed to sample knowledge peculiar to any particular group. Rather, they were intended to recognize the contribution to American culture of people who happened to be black (in the case of black items) or of people who happened to live in the last few decades (in the case of modern items). Items classified as black items dealt with aspects of the experience and culture contributed by black authors, artists, writers, musicians, and important historical and political black figures. Items classified as modern items dealt with aspects of contemporary culture and contemporary affairs such as recent developments in music, art, literature, social philosophy, social change, and politics that are not oriented toward the black culture. Items classified as traditional items dealt with such topics as world trade, geographical aspects of countries, chronological development of various historical events, differences between political systems, artifacts of various cultures, authors and aspects of world literature, music, and dance but which did not emphasize aspects of either black or modern culture. The definitions of these categories were purposely kept rather general in order to provide a conservative test of the hypothesis concerning the differences between different types of items. It was intended that, like the traditional items, these items would sample knowledge that any well-informed candidate should possess. Every item underwent the usual procedures of construction and review and met the standards of quality customarily maintained for NTE items.

PROCEDURE

It is convenient to think of the study as made up of two separate experiments. The original study used Form A and was conducted in 1970; the replication study used both Forms A and B and was conducted one year later, in 1971, and constituted in effect a replication of the first study using new candidates and items.

The Sample of Candidates for the Original Study

For the original study, a list of the centers in which the NTE was to be administered on the regular testing date was obtained, together with the record of the number of candidates tested in each of these centers one year earlier. Centers were classified as urban if they were located in cities of 100,000 or more, or as rural if they were located in cities of less than 100,000 (Long, 1968). Each center was further classified as predominantly black or not according to McGrath (1965). The distribution of the four blocks of available centers is given in Table 2.

Insert Table 2 about here

The centers within each of these four blocks were then arranged sequentially by their center code numbers, the blocks were arranged in a pre-determined sequence, and then a cumulative frequency range of candidates across blocks was computed based on the number of candidates who took the NTE at that center a year earlier.

A total of 36,426 candidates had taken the NTE in these centers one year earlier. The original research plan called for sampling approximately 2,800 candidates implying that approximately 13 centers should be sampled if the candidates were equally distributed across testing centers. The first random number selected was required to be less than the total number of candidates tested in the first block, the black-rural block, in order to insure that at least one black-rural center would be selected. Once this first random number was chosen, the number 2,802 (36,426 ÷ 13 = 2,802) was added to that first number 12 separate times as a form of cluster sampling, and the 13 test centers which contained a cumulative frequency matched by this resulting number set were selected to be included in the study. If a center was selected which had fewer than 50 candidates taking the test one year earlier, then both that center and the center adjacent to it within the block which contained the smaller number of candidates of the two adjacent centers were selected. This increased the number of selected centers from 13 to 16.

In order to reduce the logistical problems involved in using 16 centers, the list of 16 centers was reduced to eight centers. Of the original 16 centers, the one black-rural center and the two black-urban centers were retained; the three white-urban centers were reduced to two; and the 10 white-rural centers were reduced to three. One white-urban center found it impossible to cooperate on the collection of racial data, one white-urban center tested only 25 college seniors, and one white-rural center was mailed the wrong test form, thus reducing the number of test centers to five. The final list of selected centers is shown in Table 3 together with the number of seniors of each type in each center. In Table 3, population in the first column refers to the size

of the population in the city in which the NTE testing center was located, as given in Long (1968).

Insert Table 3 about here

In each of the remaining five centers, one or more groups of 44 college seniors of the same race and sex were drawn at random; the seven groups obtained constitute the sample of candidates studied, and these are designated in Table 3.

The Sample of Candidates for the Replication Study

In order to examine more closely the pattern of test results found in Form A of the NTE Common Examinations, two additional studies were carried out. The first study involved repeating the identical, critical test items used in Form A while the second study involved a different set of items in Form B; the data for both additional studies were collected at the same time one year after the original data for Form A were collected.

The test items for Form B met the same specifications and review criteria as those of Form A, and the exact same number of items of each type occurred in Form B as had occurred in Form A (Table 1); Form B had 10 items in the Traditional-Literature subtest, and five of these items were randomly selected to be used in the replication study.

The centers for the replication studies were selected in the following manner. All of the centers with at least 300 registered candidates one year earlier than the scheduled replication testing date were grouped into a set; there were four centers in this set, all in Texas, and two centers were randomly chosen. Forms A and B were mixed together in these two centers in order to permit a comparison of the two forms of the test for candidates within the same testing center.

Next, the states of the United States were grouped into five regions defined by Coleman et al. (1966, p. 39). All centers with at least 140 registered candidates one year earlier were grouped into the Northeastern and Midwestern regions; the eight such Northeastern centers were paired so that centers in cities with approximately the same population size (Long, 1968) and in close geographical distance were matched. From the Northeastern pairs, one matched pair was selected without replacement and this pair happened to include two centers in the same city in New York State; Form A was randomly assigned to one of these centers while Form B was assigned to the other center. A backup center in a large city in Pennsylvania was arbitrarily selected to provide an additional replication center, and Form A was assigned to this center.

The four Midwestern centers were located such that three of the centers were in the same city in Illinois and the one other center was located in a different state and in a city in which the population was five times less than the city in which the other three centers were located, and therefore two centers were randomly selected without replacement from the three centers that were in the same city, with one of these centers randomly assigned to Form A and the other center assigned to Form B.

Next, the black-urban and white-urban centers which had at least 140 registered candidates one year earlier were grouped into sets. Each black-urban center was paired with a white-urban center so that the pairs were as close in geographical distance and population size as possible. One black-urban center was randomly paired with one of the two white-urban centers which were located in the same city, two additional matched pairs were formed which had centers in the same city, and the last matched pair was formed by pairing the remaining black-urban center with the closest center in geographical

distance and population size of the three remaining white-urban centers. Of these four matched-pairs, two were randomly selected, and one of these pairs was randomly assigned to Form A while the other pair was assigned to Form B; as it turned out, both centers in each of these pairs were located in the same city.

Next, the two black-rural centers with at least 140 registered candidates one year earlier were each paired with a white-rural center of comparable population in the same state. One pair was randomly assigned to Form A while the other pair was assigned to Form B. Unfortunately, the white-rural center assigned to Form A had merged with another college since McGrath's (1965) classification, and all of the candidates tested in that center were black instead of white, so the planned black-rural vs. white-rural comparison could not be completed for Form A; instead, the black-rural center in Mississippi was compared to the white-rural center in Maryland, thus confounding geographical region. In addition, that white-rural center in Maryland was compared to the center in Mississippi that, in fact, turned out to be a black-rural center instead of a white-rural center.

Finally, the white-rural centers in the Northeast which had at least 140 registered candidates one year earlier were grouped into a set. Four centers resulted, of which one was arbitrarily deleted since it was surrounded by a dense population; an additional center was arbitrarily set aside as a back-up center because its population was considerably larger than the two remaining centers. Of the two remaining centers, one was randomly assigned to Form A while the other was assigned to Form B. Unfortunately, the center assigned to Form B could not cooperate in data collection, and the placement center, while used in the data collection, did not have a sufficiently large sample of college seniors to be used in the analysis.

As it turned out, none of the centers selected to participate in the replication studies had been used in the original study one year earlier. The number of college seniors tested in the replication studies, by subgroup, appear in Table 4 and Table 5. In Tables 4 and 5, population in the first column refers to the size of the population in the city in which the NTE testing center was located, as given in Long (1968).

Insert Tables 4 and 5 about here

In each of the centers which participated in the original study, as well as the centers which participated in the replication study, an observer unobtrusively coded the candidate's race as black, white, or other.

DATA ANALYSIS PROCEDURES

The basic data analysis procedure used in these three studies was an analysis of variance design in which two factors were considered fixed (Groups and Forms) and two factors were considered random (Candidates and Items) such that each random factor nested within one of the fixed factors. The number of items answered correctly within each subtest was taken as the basic score unit.

All analyses began with the same design, which was basically a candidates by items analysis of variance (cf. Hoyt, 1941). The between-candidates variance was partitioned into two portions, one of which compared performances of candidates in different groups while the other was based on differences between candidates within the same homogeneous group.

The variance between items was also partitioned such that one portion with eight degrees of freedom contrasted items from different subtests among the nine subtests shown in Table 1 while the other compared items within the same subtest. Between-subtest comparisons were further partitioned into the eight contrasts given in Table 6. The contrasts were orthogonal to each other in the sense that they were based on independent bits of information. Each contrast represents a hypothesis about differences in subject matter content of items, the culture they represent, or both interacting with each other.

Insert Table 6 about here

Table 7 shows the basic design of the analysis without the partitioning of the eight degrees of freedom just described, assuming that groups and forms are "fixed effects" and that candidates within groups and items within forms are random ones. The two parameters of primary interest are θ^2_{GF} and σ^2_{CF}, and the two hypotheses of concern are:

$$H_1: \quad \theta^2_{GF} = 0$$

$$H_2: \quad \sigma^2_{CF} = 0$$

Rejection of H_1 implies that the difference between group means differs on different forms of the test. Rejection of H_2 means that candidates within the same group rank in different orders on different forms of the test.

Insert Table 7 about here

It is apparent that if H_2 were true, the mean squares obtained in lines 7 and 8 of Table 7 should differ only by chance, since their expected values would be the same. The F ratio comparing these two mean squares may therefore be used to test the hypothesis.

A test of H_1 is not so easily derived. If H_2 is accepted, the mean squares in lines 5 and 6 of Table 7 will have the same expected value and can be used to test H_1 . If H_2 is rejected, the hypothesis that $\sigma^2_{GI} = 0$ may be tested by comparing the mean squares in lines 6 and 8 of Table 7. If this hypothesis is accepted, H_1 may be tested by comparing the mean squares in lines 5 and 7; otherwise, assuming H_2 to be false, and approximate procedure (Winer, 1962, pp. 201-202) is available.

If the effect of forms is regarded as fixed then each of the eight mean squares of the between forms comparison has an expected value of the form shown in Table 7 for line 3 except that the parameter θ^2_F will have a different value. The same thing is true for the Groups X Forms interaction and for the Candidates X Forms interaction. Rejection of H_1 in any given instance implies that group means differ more on items of one of the two types contrasted in that line, and results in the Candidates X Forms interaction are similarly interpretable in terms of item types. In other words, both H_1 and H_2 were tested separately with regard to eight alternatives to the null hypothesis.

Before a final decision was made with regard to any of these hypotheses, however, the analysis was simplified. In every instance in which an hypothesis about any parameter was testable, the test was made, and if the null hypothesis was acceptable with an F smaller than 2.00 (to minimize the risk of accepting false hypotheses), the corresponding sum of squares were pooled with their respective error mean squares (Green & Tukey, 1960).

Since results of these last two types, relevant to H_1 or H_2 , are the ones of primary concern, they are the only ones that are examined in detail within this article.

RESULTS

The Original Study

The strategy adopted in analyzing the data for the original study was to conduct four separate analyses, each using one pair of groups chosen to differ on a variable hypothesized to relate to test performance. Three pairs of groups which differed in race were compared in the original study: B and E, A and G, and C and D. Unfortunately, the racial variable was confounded in each instance with some other variable. Contrasts between B and E, and also C and D, both used black candidates tested in an urban location (140,000) and white candidates tested in a rural one (10,000) in Mississippi, so that urban-rural factors could distort the results. Groups A and G compared black women in Louisiana with white women in West Virginia, and here also the black candidates were examined in a larger city than the whites. In this latter instance either a geographical or an urban- "semi-urban" influence could distort the results. Groups C and F (Table 3) were compared because they differed with respect to size of the community in which they were tested, but did not differ in race or sex, so that urban-rural differences (if any) might be detected.

The relevant findings from all four analyses of variance are presented in Table 8 and summarized in Table 9. The interaction between candidates and item types is of interest because it may be interpreted as indicating that candidates within the same group are ranked in different orders by items of the two types; that is, that the individuals who score the highest on one type of item are not the same ones as score highest on the other type. In short, a test made up entirely of one type of item would measure a different function than one made up entirely of items of the other type.

Insert Tables 8 and 9 about here

Notice, for example, that in the comparison of Groups A and G, candidates interact significantly with social studies items vs. literature and art items (Table 9, line 6), implying that a particular candidate would not do equally well on both a social studies and a literature-art test. This is not surprising, since such tests would sample knowledge in two quite distinct fields of knowledge. Rather, it seems odd that in the other two pairs of different racial groups this was not the case. This may be taken as indicating that the correlation between two tests like these would be substantial enough so that any differences in what they measure would be difficult to detect when only 30 items of each type are used.

Differences in the dimensions measured by literature-art items are even more difficult to detect (Table 9, line 5). In none of the four pairs of groups was any evidence found that art knowledge was any different from literary knowledge. It should be noted that only 15 items of each type were used, however.

Differences between a test made up of traditional items and one made up of equal numbers of black and modern items (Table 9, line 4) were also difficult to distinguish. (Since the statistical test here is based on comparing intraclass correlations between the two tests with the intraclass correlations within them, it is likely that a rather low intercorrelation between black and modern items has a lot to do with this.)

In contrast, significant interactions between candidates and black vs. modern items were found in three of the four analyses (Table 9, line 3). This suggests that culture represented is probably a more important factor in a candidate's performance when he is compared with others of his own race, sex, and institutional affiliation than is subject-matter content.

Data on individual candidates are not available to indicate how black candidates who know more about black or modern culture than about traditional culture differ from black candidates of the same sex (tested in the same center) who know less about black or modern culture than they do about traditional culture.

Of most interest are interactions between groups and item types because they indicate group differences in mean performance on the different types of items. No group differences related to subject-matter content were found which suggests that the racial groups differ by about the same amount on the different subject matter areas. Differences between black and modern items, on the other hand, were found in all three racial comparisons. Differences between black and modern items combined and traditional items were also found in all three racial contrasts, as well as in the one between black women in urban Mississippi and in rural Virginia.

To facilitate comparison of results obtained with different pairs of groups, Table 10 shows the differences between groups in each pair. The advantage to black candidates of shifting from traditional to half black and half modern items in each instance approaches 20% of the total number of items. The advantage to black candidates that would result from a change from all modern to all black items would be about the same. Finally, a change from all traditional to all black items would boost the score of the average black candidate relative to that of the average white candidate by almost 19 points, nearly 30%!

Insert Table 10 about here

The comparison of female black candidates from urban Mississippi and ones from rural Virginia (Groups C and F) indicated that the two groups differed much more on traditional items than on black and modern ones. The rural Virginia women were distinctly superior on traditional items and only slightly better on black-modern ones, so that the urban Mississippi group would benefit by almost 11 points if the comparison were made on black-modern items rather than on traditional ones.

Replication of Form A

The relevant findings from the 1971 replication using Form A are presented in Table 11 and summarized in the seven analyses of variance given in Table 12. A significant interaction between candidates in the same racial group and the difference between literature and art items occurs in only 1 of the 7 racial comparisons (Table 12, line 8). Except in this case, no difference between the

candidates' knowledge of art or literature was found when 15 items of each type were used. Similarly, a significant second-order interaction (candidate by traditional vs. social studies by traditional vs. black and modern items) occurs in only 1 of the 7 racial comparisons. No attempt will be made to interpret this finding here.

Insert Tables 11 and 12 about here

In contrast, candidates interact significantly with social studies vs. literature and art items (Table 12, line 9) in 6 of the 7 racial comparisons implying that not all candidates would do equally well on both a social studies and a literature and art test, confirming the fact that the two tests sample distinct fields of knowledge.

Differences between a test made up of traditional items and one made up of equal numbers of black and modern items (Table 12, line 7) were significant in 3 of the 7 racial comparisons.

In contrast, significant interactions between candidates and black vs. modern items were found in 5 of the 7 racial comparisons (Table 12, line 6). This suggests that when candidates of the same sex and race are compared, the culture represented by the items is an important factor in determining their relative achievement, along with subject-matter content. Groups A and B, composed of urban females in North Carolina, and Groups M and N, composed of males in rural Mississippi and Maryland, did not show this phenomenon.

Interactions between groups and item types indicate group differences in mean performance on the different types of items and are therefore more closely relevant to this discussion. Group differences related to subject-matter content in second-order interactions involving cultural origins were found in two instances (Table 12, lines 4 and 5) and in two pairs of groups in which groups interacted with social studies vs. the literature and art items. As in the first study, little evidence was found to suggest that group differences were greater in any one subject area than in another.

Differences between black and modern items occurred in 6 of the 7 racial comparisons, however, and the seventh comparison was very close to significance; similar results were obtained with respect to black and modern items combined vs. traditional items.

Table 13 shows the differences between groups in each pair. The advantage to black candidates of shifting from traditional to half black and half modern items in the 6 of the 7 racial comparisons that show significant differences averaged 11 points on a test of a projected length of 65 items. The advantage to black candidates that would result from a change from all modern to all black items would be almost 12 points. Finally, a change from all traditional to all black items would boost the score of the average black candidate relative to that of the average white candidate by an average of 16 points, nearly 25%. Except in the two cases in which Mississippi blacks were compared with Maryland whites, the results are highly consistent with those obtained in the first study (cf. Table 10).

Insert Table 13 about here

Form B

The relevant findings from the 1971 administration of Form B are presented in Table 14 and summarized in Table 15. Only 1 of the 4 racial comparisons showed a significant interaction for each of the following: candidates by black vs. modern items (Table 15, line 5), candidates by traditional items vs. equal numbers of black and modern items combined (Table 15, line 6), and candidates by literature vs. art items (Table 15, line 7). Since these results were not consistent across groups, they will not be discussed any further here.

Insert Tables 14 and 15 about here

Two of the four racial comparisons showed a significant second-order interaction for the candidates by traditional vs. black-modern by literature vs. fine-art items and for candidates by traditional vs. black-modern by social studies vs. literature and art items.

Candidates interacted significantly with social studies vs. literature and art items (Table 15, line 8) in 3 of the 4 racial comparisons; this suggests that social studies items do indeed sample a different field of knowledge than that sampled by literature and art items.

Interactions between groups and item types, which indicate group differences in mean performance on the different types of items, related to subject-matter content occurred in two analyses--one of which involved a second-order interaction.

Interactions between race and differences between black and modern items, and between race and differences between black and modern items combined and traditional items were found in all four comparisons. Table 16 shows the differences between the groups in each pair. The advantage to black candidates of shifting from traditional to half black and half modern items was significant in all four racial comparisons, with an average advantage of almost 10 points on a test of a projected length of 65 items. The advantage to black candidates that would result from a change from all modern to all black items would be almost 22 points. (The median difference between black and modern items was larger for Form B than for Form A--22 points vs. 12 points.) Finally, a change from all traditional to all black items would boost the score of the average black candidate relative to that of the average white candidate by an average of 20 points, more than 30%. Once again, strikingly consistent results were obtained.

Insert Table 16 about here

SUMMARY AND DISCUSSION

The results about the impact on relative performances of black and white candidates that relate to decisions about relative numbers of black, modern, and traditional items included in the NTE are quite consistent. During the past two years the content specifications of the social studies, literature and fine arts subtest of the NTE Common Examinations have been changed to include a larger proportion of items that we have classified as black or modern, and these changes in the test specifications reflect both the influence of a concern for the recognized contribution of these types of items and the contribution of research along the lines described in this article.

Fortunately, the contribution to total test score made by this subtest is small. The description of the content validity of the NTE would involve a study not only of the curricular offerings of teacher training programs but also of an analysis of the types of knowledge needed by teachers in the classrooms. In this research study, we accepted the content of the NTE subtest that we studied as reasonable, and concerned ourselves with the influence of cultural emphases within these content areas.

While it is true the comparisons made between the races in the original study were in each instance confounded with differences in where the test was taken, which may have allowed regional or urban-rural differences to distort the racial comparisons, this is not true of the replication study done in 1971, and yet the results were virtually identical in the latter study. These results leave little doubt that black candidates tend to possess a different set of knowledge than white candidates, and that these differences have little to do with conventional subject-matter areas. The differences reflect differences in the particular subculture sampled by the items. Black candidates know relatively more about black culture than they do about white culture.

The consistency of the findings across the three studies is striking. In all, 14 pairs of black vs. white candidates were compared in various locations. Significant interactions between groups and cultural origin of items were found in 13 of the 14 pairs. The median advantage to black candidates over white candidates if a projected test consisting of 65 black items were used instead of a projected test of 65 modern items would be 12 points.

When black and modern items combined were compared with traditional items, 13 of the 14 comparisons produced significant differences, and the median advantage to black candidates was estimated to be 11 points if a test of half black and half modern items were used instead of a projected test on which all 65 items were traditional.

The contrast between the black and traditional items was not tested directly (because it was not orthogonal to the other contrasts tested); however, the median advantage to black candidates over white candidates if a projected test of 65 black items were used instead of a projected test of 65 traditional ones was estimated to be 18 points--almost 30%.

The most interesting question left unanswered in this study is, if within a group homogeneous as to sex, race, and institution in which tested, some candidates do best on black items, some do best on modern ones, and some do best on traditional ones, in what other ways are these three types of candidates different? Does this "selective learning factor" reflect differences in interests or concerns of individual candidates which might relate to their effectiveness as teachers in urban schools vs. suburban schools, or in integrated schools vs. segregated ones?

REFERENCES

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. Equality of educational opportunity. Washington, D. C.: U. S. Government Printing Office, 1966.

Green, B. F., Jr., & Tukey, J. W. Complex analyses of variance: General problems. Psychometrika, 1960, 25, 127-152.

Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.

Long, L. H. The world almanac and book of facts. New York: Newspaper Enterprise Association, Inc., 1968.

McGrath, E. J. The predominantly Negro colleges and universities in transition. New York: Bureau of Publications, Teachers College, Columbia University, 1965.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

FOOTNOTES

[1]This project was supported by the National Teacher Examinations Program at Educational Testing Service. Several colleagues at Educational Testing Service have provided valuable assistance during this project; in particular, we would like to thank Richard Majetic for his patience and encouragement, and Kitty Kornhauser and Stan MacDonald for writing the necessary computer program for the analyses. We would especially like to acknowledge the contribution of the Test Development Staff at Educational Testing Service who provided excellent cooperation with us both in terms of providing the necessary types of items through the usual procedures and in meeting our demanding schedule requirements.

[2]In Tables 3, 4, 5 "population" refers to the size of the population in the city in which the NTE testing center was located, as given in Long (1968).

[3]All means in Tables 10, 13, 16 are based on projected tests of 65 items of the type indicated.

Table 1

Item Distribution for the Experimental Subtests

NTE Common Examination Subtests	Item Type			Total
	Black	Traditional	Modern	
Fine Arts	5	5	5	15
Literature	5	5	5	15
Social Studies	10	10	10	30
Total	20	20	20	60

Table 2

Frequency of Four Types of NTE Centers

Type of Center	Original Study: Number Available	Replication Study: Number Available with at Least 140 Candidates 1 Year Earlier
Black-Urban	16	4
Black-Rural	21	2
White-Urban	103	38
White-Rural	221	44
Total	361	88

Table 3

Composition of the Samples of College Seniors Used in the Original Study

Population[2]	State	Candidates				Sample Selected*	
		Male		Female		Designation	Type
		White	Black	White	Black		
150,000	Louisiana	0	11	0	67	A	black female
140,000	Mississippi	0	49			B	black male
				0	84	C	black female
10,000	Mississippi			113	2	D	white female
		44	0			E	white male
40,000	Virginia	0	26	1	100	F	black female
90,000	West Virginia	23	0	101	1	G	white female
	Totals	67	86	215	254		
		153		469			
		622					

*Each of the seven samples (A through G) contained 44 candidates randomly selected.

Table 4

Composition of the Samples of College Seniors Used in the Replication of Form A

Population[2]	State	Candidates				Sample Selected		
		Male		Female				
		White	Black	White	Black	Designation	Type	N*
7,800,000	New York	65	4	264	35	G	white female	34
						H	black female	34
3,600,000	Illinois	11	0	60	3			
130,000	North Carolina	0	14	1	56	A	black female	44
130,000	North Carolina	48	0	241	10	B	white female	44
Less than 2500	Mississippi	0	42	0	79	C	black female	44
						I	black female	34 (selected randomly from C)
						J	black male	34
Less than 2500	Mississippi	0	20	0	64	L	black female	44
						N	black male	20
7,000	Maryland	92	0	143	0	D	white female	44
						K	white male	34
						M	white male	20 (selected randomly from K)
2,000,000	Pennsylvania	49	4	128	34	E	white female	34
						F	black female	34

*Sample size selected randomly from the total subgroups.

Table 5

Composition of the Samples of College Seniors Used in Form B

Population[2]	State	Candidates				Sample Selected		
		Male		Female				
		White	Black	White	Black	Designation	Type	N*
7,800,000	New York	32	2	188	6			
3,600,000	Illinois	40	4	92	21	F	white female	21
						G	black female	21
900,000	Texas	3	9	3	50	B	black female	44
900,000	Texas	6	1	64	3	A	white female	44
Less than 2500	Alabama	0	18	0	104	C	black female	44
Less than 2500	Alabama	17	0	113	1	D	white female	44
90,000	Massachusetts	4	0	18	0			
900,000	Texas	28	1	132	4	E	white female	44

*Sample size selected randomly from the total subgroups.

Table 6

Eight Contrasts Between Item Types

Name of Contrast	Form 1	Form 2	Total Number of Items Involved
1. Black	20 Black items	20 Modern items	40
2. Traditional	20 Black items plus 20 Modern items	20 Traditional items	60
3. Literature	15 Art items	15 Literature items	30
4. Social Studies	15 Art items plus 15 Literature items	30 Social Studies items	60
5. Black X Literature (B X L)	5 Black Art items plus 5 Modern Literature items	5 Modern Art items plus 5 Black Literature items	20
6. Traditional X Literature (T X L)	5 Black Art items plus 5 Modern Art items plus 5 Traditional Literature items	5 Traditional Art items plus 5 Black Literature items plus 5 Modern Literature items	30
7. Black X Social Studies (B X S)	5 Black Art items plus 5 Black Literature items plus 10 Modern Social Studies items	5 Modern Art items plus 5 Modern Literature items plus 10 Black Social Studies items	40
8. Traditional X Social Studies (T X S)	5 Black Art items plus 5 Black Literature items plus 5 Modern Art items plus 5 Modern Literature items plus 10 Traditional Social Studies items	5 Traditional Art items plus 5 Traditional Literature items plus 10 Black Social Studies items plus 10 Modern Social Studies items	60

Table 7

Expected Mean Squares for Basic Analysis of Variance*

Source of Variation	df	Expected Mean Square
1. Between Groups	1	$2{,}640\,\theta^2_G + 60\,\sigma^2_C + 44\,\sigma^2_{GI} + \sigma^2$
2. Candidates (Within Groups)	86	$60\,\sigma^2_C + \sigma^2$
3. Between Forms	8	$660\,\theta^2_F + 88\,\sigma^2_I + 7.5\,\sigma^2_{CF} + \sigma^2$
4. Items (Within Forms)	51	$88\,\sigma^2_I + \sigma^2$
5. Interaction, Groups X Forms	8	$330\,\theta^2_{GF} + 44\,\sigma^2_{GI} + 7.5\,\sigma^2_{CF} + \sigma^2$
6. Interaction, Groups X Items	51	$44\,\sigma^2_{GI} + \sigma^2$
7. Interaction, Candidates X Forms	688	$7.5\,\sigma^2_{CF} + \sigma^2$
8. Residual Variation	4,386	σ^2
9. Total Variation	5,279	

*When N = 44 candidates per group.

Table 8

Simplified Analysis of Scores of Candidates of the First Study

Source of Variation	Groups															
	A vs. G				C vs. D				B vs. E				C vs. F			
	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F
1. Between Groups	1	4.38	4.38*	4.33	1	0.59	0.59(n.s.)	< 1	1	1.04	1.04(n.s.)	1.22	1	11.37	11.37**	10.39
2. Candidates (Within Groups)	86	56.22	0.65**	3.82	86	48.80	0.57**	3.60	86	49.05	0.57**	3.35	86	46.98	0.55**	3.44
3a. Black vs. Modern Items	1	23.89	23.89**	7.20	1	19.35	19.35**	7.77	1	11.94	11.94*	4.73	1	50.83	50.83**	14.95
3b. Traditional vs. Black and Modern Items	1	7.75	7.75(n.s.)	2.33	1	1.13	1.13(n.s.)	< 1	--	-----	-----	----	--	-----	-----	-----
3c. Literature vs. Art Items	--	----	-----	----	1	11.08	11.08*	4.86	1	10.31	10.31*	4.46	1	11.60	11.60 (n.s.)	3.65
3d. Social Studies vs. Art and Literature Items	1	0.27	0.27(n.s.)	< 1	--	-----	-----	----	1	7.43	7.43(n.s.)	3.22	1	3.15	3.15 (n.s.)	< 1
3f. Traditional X Literature Items	1	18.33	18.33*	5.52	1	10.06	10.06*	4.41	1	10.06	10.06*	4.35	1	9.82	9.82 (n.s.)	3.09
3g. Black X Social Studies Items	1	10.47	10.47(n.s.)	3.15	1	9.72	9.72*	4.26	1	6.83	6.83(n.s.)	2.96	--	----	----	----
4. Items Within Forms	54	182.41	3.38**	19.88	54	123.27	2.28**	14.25	54	124.69	2.31**	13.59	55	174.74	3.18**	19.88
5a. Groups X Black vs. Modern Items	1	7.27	7.27**	18.18	1	10.15	10.15**	13.93	1	7.37	7.37**	11.25	1	0.21	0.21 (n.s.)	< 1
5b. Groups X Traditional vs. Black and Modern Items	1	12.55	12.55**	31.38	1	11.14	11.14**	15.50	1	8.93	8.93**	21.26	1	7.91	7.91**	14.13
5g. Groups X Social Studies vs. Art and Literature Items	--	-----	-------	-----	--	-----	-------	-----	--	----	-------	-----	1	0.71	0.71 (n.s.)	1.12
5h. Groups X Traditional X Social Studies Items	--	-----	-------	-----	--	-----	-------	-----	1	1.67	1.67(n.s.)	3.98	--	-----	------	-----
6. Groups X Items Within Forms	57	22.54	0.40**	2.35	57	29.19	0.51**	3.19	56	23.32	0.42**	2.47	56	31.47	0.56**	3.50
7a. Candidates X Black vs. Modern Items	--	-----	------	----	86	19.52	0.23**	1.44	86	21.12	0.25**	1.47	86	19.84	0.23**	1.44
7b. Candidates X Traditional vs. Black and Modern Items	--	-----	------	----	86	18.81	0.22*	1.38	--	-----	------	----	--	-----	------	-----
7d. Candidates X Social Studies vs. Art and Literature Items	86	19.79	0.23*	1.35	--	-----	-----	-----	--	-----	------	----	86	18.96	0.22**	1.38
8. Residual Variation	4,988	835,93	0.17		4,902	779.69	0.16		4,988	852.86	0.17		4,902	776.61	0.16	
9. Total Variation	5,279	1201.80			5,279	1092.50			5,279	1136.62			5,279	1164.20		

* .01 < P < .05
** P < .01
n.s. P > .05

Table 9

Summary of Findings with Respect to Group and Candidate Differences in the Original Study

Source of Variation	Groups Compared			
	A vs. G Females- Louisiana Black vs. West Virginia White	C vs. D Mississippi Females- Urban Black vs. Rural White	B vs. E Mississippi Males- Urban Black vs. Rural White	C vs. F Black Females- Urban Mississippi vs. Rural Virginia
Interactions, Groups by Item Types:				
(1) Groups by (Black vs. Modern Items)	**	**	**	n.s.
(2) Groups by (Traditional vs. Black and Modern Items)	**	**	**	**
Interactions, Candidates in Same Group by Item Types:				
(3) Candidates by (Black vs. Modern Items)	n.s.	**	**	**
(4) Candidates by (Traditional vs. Black and Modern Items)	n.s.	*	n.s.	n.s.
(5) Candidates by (Literature vs. Art Items)	n.s.	n.s.	n.s.	n.s.
(6) Candidates by (Social Studies vs. Literature and Art Items)	*	n.s.	n.s.	**

* .01 < P < .05 (significant, .05 level)
** P < .01 (significant, .01 level)
n.s. P > .05 (not significant, .05 level)

Table 10

Significant Group Differences Related to Types of Items[3] in the Original Study

Composition of Test	Differences Between Group Mean Numbers Right			
	A vs. G Females: Louisiana Black vs. West Virginia White	C vs. D Mississippi Females: Urban Black vs. Rural White	B vs. E Mississippi Males: Urban Black vs. Rural White	C vs. F Black Females: Urban Mississippi vs. Rural Virginia
65 Black Items	+6.6	+9.9	+7.9	--
65 Modern Items	-5.2	-4.2	-3.9	--
Difference, B-M	+11.8**	+14.1**	+11.8**	n.s.
65 Items, Half Black, Half Modern	+0.8	+2.9	+1.9	-2.4
65 Traditional Items	-12.7	-9.9	-9.4	-13.2
Difference, (B+M)-T	+13.5**	+12.8**	+11.3**	+10.8**
Difference, B-T (Not tested)	+19.3	+19.8	+17.3	

* $.01 < P < .05$
** $P < .01$
n.s. $P > .05$

Table 11

Simplified Analysis of Scores of Candidates in the Replication of Form A

Source of Variation	Groups															
	J vs. M				G vs. H				E vs. F				B vs. L			
	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df
1. Between Groups	1	12.18	12.18*	9.50	1	0.57	0.57(n.s.)	<1.0	1	0.97	0.97(n.s.)	<1.0	1	17.27	17.27*	14.30	1
2. Candidates (Within Groups)	38	33.15	0.87*	5.12	66	52.24	0.79*	4.51	66	73.51	1.11*	6.38	86	50.77	0.59*	3.41	86
3a. Black vs. Modern Items	—	———	———	———	1	33.93	33.93*	9.21	1	24.85	24.85*	8.23	1	14.00	14.00*	4.60	1
3b. Traditional vs. Black and Modern Items	—	———	———	———	1	1.54	1.54(n.s.)	<1.0	1	5.51	5.51(n.s.)	1.86	—	———	———	———	—
3c. Literature vs. Art Items	1	8.33	8.33*	5.37	—	———	———	———	1	22.66	22.66*	8.09	—	———	———	———	1
3d. Social Studies vs. Art and Literature Items	1	3.92	3.92(n.s.)	2.25	1	0.001	0.001(n.s.)	<1.0	1	0.00	0.00(n.s.)	<1.0	1	8.35	8.35(n.s.)	2.71	1
3f. Traditional X Literature Items	1	6.75	6.75*	4.35	—	———	———	———	—	———	———	———	—	———	———	———	—
3g. Black X Social Studies Items	—	———	———	———	—	———	———	———	—	———	———	———	—	———	———	———	1
3h. Traditional X Social Studies Items	—	———	———	———	—	———	———	———	—	———	———	———	—	———	———	———	—
4. Items Within Forms	56	87.04	1.55*	9.13	56	187.39	3.35*	19.14	55	153.82	2.80*	16.09	57	161.47	2.83*	16.36	55
5a. Groups X Black vs. Modern Items	1	2.03	2.03*	4.72	1	5.84	5.84*	8.99	1	5.65	5.65*	9.39	1	3.32	3.32*	3.97	1
5b. Groups X Traditional vs. Black and Modern Items	—	———	———	———	1	6.71	6.71*	11.11	1	3.97	3.97*	6.57	1	9.46	9.46*	13.02	1
5d. Groups X Social Studies vs. Art and Literature Items	—	———	———	———	—	———	———	———	—	———	———	———	—	———	———	———	1
5e. Groups X Black X Literature Items	—	———	———	———	1	2.65	2.65*	8.55	—	———	———	———	—	———	———	———	—
5h. Groups X Traditional X Social Studies Items	—	———	———	———	—	———	———	———	1	1.69	1.69*	4.45	—	———	———	———	—
6. Groups X Items Within Forms	58	25.01	0.43*	2.53	56	17.59	0.31*	1.77	58	21.37	0.38*	2.18	57	36.02	0.63*	3.64	56
7a. Candidates X Black vs. Modern Items	—	———	———	———	66	23.79	0.36*	2.07	66	13.64	0.24*	1.37	86	21.89	0.25*	1.45	86
7b. Candidates X Traditional vs. Black and Modern Items	—	———	———	———	66	20.75	0.31*	1.77	66	16.84	0.25*	1.44	—	———	———	———	—
7c. Candidates X Literature vs. Art Items	—	———	———	———	—	———	———	———	—	———	———	———	—	———	———	———	—
7d. Candidates X Social Studies vs. Art and Literature Items	38	10.08	0.27*	1.59	66	19.24	0.29*	1.64	66	16.75	0.25*	1.44	86	26.81	0.31*	1.80	86
7h. Candidates X Traditional X Social Studies Items	—	———	———	———	—	———	———	———	—	———	———	———	—	———	———	———	—
8. Residual Variation	2,204	376.22	0.17		3,696	647.31	0.175		3,696	642.96	0.174		4,902	848.23	0.173		4,902
9. Total Variation	2,399	564.71			4,079	1,019.82			4,079	1005.99			5,279	1197.59			5,279

* P<.05

Table 11

Simplified Analysis of Scores of Candidates in the Replication of Form A

Groups	E vs. F				D vs. L				C vs. D				A vs. B				J vs. K			
	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F
F	1	0.97	0.97(n.s.)	<1.0	1	17.27	17.27*	14.30	1	12.90	12.90*	9.98	1	14.43	14.43*	9.36	1	37.86	37.86*	25.67
4.51	66	73.51	1.11*	6.38	86	50.77	0.59*	3.41	86	58.62	0.68*	3.89	86	72.11	0.84*	4.94	66	64.06	0.97*	5.74
9.23	1	24.85	24.85*	8.23	1	14.00	14.00*	4.60	1	22.43	22.43*	8.73	1	31.88	31.88*	11.47	1	12.18	12.18*	4.70
.)<1.0	1	5.51	5.51(n.s.)	1.86	—	-----	-----	-----	—	-----	-----	----	—	-----	-----	----	1	1.77(n.s.)	1.77(n.s.)	<1.0
----	1	22.66	22.66*	8.09	—	-----	----	-----	1	10.95	10.95*	4.68	1	6.01	6.01(n.s.)	2.04	—	-----	-----	-----
.)<1.0	1	0.00	0.00(n.s.)	<1.0	1	8.35	8.35(n.s.)	2.71	1	1.64	1.64(n.s.)	<1.0	1	0.17	0.17(n.s.)	<1.0	—	-----	-----	-----
---	—	-----	-----	----	—	-----	-----	----	—	-----	-----	----	—	-----	-----	----	—	-----	-----	-----
---	—	----	----	----	—	-----	-----	----	1	9.31	9.31(n.s.)	3.98	—	-----	-----	----	1	13.27	13.27*	5.35
---	—	----	----	----	—	-----	-----	----	—	-----	-----	----	1	0.11	0.11(n.s.)	<1.0	—	-----	-----	-----
19.14	55	153.82	2.80*	16.09	57	161.47	2.83*	16.36	55	128.97	2.34*	13.37	55	152.71	2.78*	16.35	56	133.55	2.38*	14.08
8.99	1	5.65	5.65*	9.39	1	3.32	3.32*	3.37	1	7.92	7.92*	9.20	1	15.42	15.42*	21.42	1	2.72	2.72(n.s.)	3.81
11.11	1	3.97	3.97*	6.57	1	9.46	9.46*	15.02	1	9.04	9.04*	14.35	1	16.00	16.00*	22.22	1	3.97	3.97*	5.67
---	—	----	----	----	—	-----	-----	-----	1	6.21	6.21*	6.95	—	-----	-----	-----	1	2.31	2.31*	4.62
8.55	—	----	----	----	—	-----	-----	-----	—	-----	-----	----	—	-----	-----	-----	—	-----	-----	-----
----	1	1.69	1.69*	4.45	—	-----	-----	-----	—	-----	-----	----	—	-----	-----	-----	—	-----	-----	-----
1.77	56	21.37	0.38*	2.18	57	36.02	0.63*	3.64	56	35.17	0.63*	3.60	57	41.04	0.72*	4.24	56	28.25	0.50*	2.96
2.07	66	13.64	0.34*	1.37	86	21.89	0.25*	1.45	86	21.12	0.29*	1.43	—	-----	-----	-----	66	16.30	0.25*	1.46
1.77	66	16.64	0.25*	1.44	—	-----	-----	-----	—	-----	-----	----	—	-----	-----	-----	66	15.27	0.23*	1.37
----	—	-----	-----	----	—	-----	-----	-----	—	-----	-----	----	86	21.52	0.25*	1.46	—	-----	-----	-----
1.66	66	16.75	0.25*	1.44	86	26.81	0.31*	1.83	86	24.74	0.29*	1.64	86	24.51	0.29*	1.71	—	-----	-----	-----
----	—	----	----	----	—	-----	-----	-----	—	-----	-----	----	86	21.16	0.25*	1.47	—	-----	-----	-----
	3,696	642.96	0.174		4,902	848.23	0.173		4,902	860.72	0.175		4,816	819.91	0.17		3,762	636.98	0.169	
	4,079	1005.99			5,279	1197.59			5,279	1209.74			5,279	1237.00			4,079	968.49		

Table 12

Summary of Findings with Respect to Group and Candidate Differences in the Replication of Form A

Source of Variation	Groups Compared						
	N vs. M Males: Mississippi Black vs. Maryland White	G vs. H New York Females: White vs. Black	E vs. F Pennsylvania Females: White vs. Black	D vs. L Females: Maryland White vs. Mississippi Black	C vs. D Females: Mississippi Black vs. Maryland White	A vs. B North Carolina Females: Black vs. White	J vs. K Males: Mississippi Black vs. Maryland White
Interactions, Groups by Item Types:							
(1) Groups by (Black vs. Modern Items)	*	*	*	*	*	*	n.s.
(2) Groups by (Traditional vs. Black and Modern Items)	n.s.	*	*	*	*	*	*
(3) Groups by (Social Studies vs. Art and Literature Items)	n.s.	n.s.	n.s.	n.s.	*	n.s.	*
(4) Groups by (B X L Items)	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.
(5) Groups by (T X S Items)	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.
Interactions, Candidates in Same Group by Item Types:							
(6) Candidates by (Black vs. Modern Items)	n.s.	*	*	*	*	n.s.	*
(7) Candidates by (Traditional vs. Black and Modern Items)	n.s.	*	*	n.s.	n.s.	n.s.	*
(8) Candidates by (Literature vs. Art Items)	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
(9) Candidates by (Social Studies vs. Art and Literature Items)	*	*	*	*	*	*	n.s.
(10) Candidates by (T X S Items)	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.

* $P < .05$ (significant, .05 level)

Table 13

Significant Group Differences Related to Cultural Reference of Items[3] on the Replication of Form A

Composition of Test	Differences Between Group Mean Numbers Right						
	N vs. M Males: Mississippi Black vs. Maryland White	H vs. G New York Females: Black vs. White	F vs. E Pennsylvania Females: Black vs. White	L vs. D Females: Mississippi Black vs. Maryland White	C vs. D Females: Mississippi Black vs. Maryland White	A vs. B North Carolina Females: Black vs. White	J vs. K Males: Mississippi Black vs. Maryland White
65 Black Items	-2.6	+11.3	+6.8	+0.4	+3.5	+6.8	-5.6
65 Modern Items	-11.9	-0.8	-5.0	-7.5	-8.7	-10.3	-13.8
Difference, B-M	+9.3*	+12.1*	+11.8*	+7.9*	+12.2*	+17.1*	+8.2 n.s.
65 Items, Half Black, Half Modern	-7.2	+5.3	+0.8	-3.6	-2.7	-1.8	-9.7
65 Traditional Items	-13.4	-5.9	-7.7	-15.3	-14.1	-16.9	-18.2
Difference, (B+M)-T	+6.2 n.s.	+11.2*	+8.5*	+11.7*	+11.4*	+15.1*	+8.5*
Difference, B-T (Not tested)	+10.8	+17.2	+14.5	+15.7	+17.6	+23.7	+12.6

* $P < .05$
n.s. $P > .05$

Table 14

Simplified Analysis of Scores of Candidates on Form B

Source of Variation	Groups															
	A vs. B				F vs. G				B vs. E				C vs. D			
	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F	df	Sum of Squares	Mean Square	F
1. Between Groups	1	55.43	55.43*	40.00	1	0.57	0.57n.s.	<1.0	1	41.13	41.13*	30.37	1	42.02	42.02*	35.45
2. Candidates (Within Groups)	86	86.12	1.00*	5.81	40	51.17	1.28*	7.53	86	74.59	0.87*	5.12	86	63.09	0.73*	4.51
3a. Black vs. Modern Items	1	15.16	15.16*	4.50	1	8.15	8.15*	4.97	1	21.33	21.33*	6.58	1	15.96	15.96n.s.	3.59
3b. Traditional vs. Black and Modern Items	—	—	—	—	—	1.12	1.12n.s.	<1.0	—	—	—	—	—	—	—	—
3c. Literature vs. Art Items	1	21.28	21.28*	6.31	1	13.41	13.41*	8.18	1	17.35	17.35*	5.35	1	24.63	24.63*	5.42
3d. Social Studies vs. Art and Literature Items	—	—	—	—	1	0.46	0.46n.s.	<1.0	1	0.78	0.78 n.s.	<1.0	1	0.16	0.16n.s.	<1.0
3f. Traditional X Literature Items	1	9.58	9.58 n.s.	2.72	—	—	—	—	1	7.86	7.86 n.s.	2.31	—	—	—	—
3g. Black X Social Studies Items	—	—	—	—	1	7.87	7.87*	4.80	1	17.47	17.47*	5.39	—	—	—	—
3h. Traditional X Social Studies Items	—	—	—	—	—	—	—	—	1	0.68	0.68 n.s.	<1.0	1	0.05	0.05n.s.	<1.0
4. Items Within Forms	55	188.50	3.37*	19.59	54	88.39	1.64*	9.65	53	171.82	3.24*	19.06	55	235.39	4.28*	26.42
5a. Groups X Black vs. Modern Items	1	21.17	21.17*	54.28	1	9.30	9.30*	29.06	1	28.37	28.37*	57.90	1	27.48	27.48*	41.25
5b. Groups X Traditional vs. Black and Modern Items	1	7.80	7.80*	20.00	1	2.36	2.36*	4.22	1	5.55	5.55*	11.33	1	6.75	6.75*	14.67
5c. Groups X Literature vs. Art Items	1	2.02	2.02*	5.18	—	—	—	—	—	—	—	—	—	—	—	—
5g. Groups X Black X Social Studies Items	—	—	—	—	—	—	—	—	1	2.30	2.30*	4.69	—	—	—	—
6. Groups X Items Within Forms	56	22.09	0.39*	2.27	57	17.96	0.32*	1.88	56	27.55	0.49*	2.88	57	26.22	0.46*	2.84
7a. Candidates X Black vs. Modern Items	—	—	—	—	—	—	—	—	—	—	—	—	86	17.99	0.21*	1.29
7b. Candidates X Traditional vs. Black and Modern Items	—	—	—	—	40	11.05	0.28*	1.65	—	—	—	—	—	—	—	—
7c. Candidates X Literature vs. Art Items	—	—	—	—	—	—	—	—	—	—	—	—	86	24.81	0.29*	1.78
7d. Candidates X Social Studies vs. Art and Literature Items	—	—	—	—	40	10.63	0.27*	1.59	86	19.35	0.23*	1.33	86	22.25	0.26*	1.60
7f. Candidates X Traditional X Literature Items	86	18.99	0.22*	1.30	—	—	—	—	86	20.72	0.24*	1.41	—	—	—	—
7h. Candidates X Traditional X Social Studies Items	—	—	—	—	—	—	—	—	86	20.56	0.24*	1.41	86	19.37	0.23*	1.42
8. Residual Variation	4,988	858.96	0.172		2,280	388.03	0.17		4,816	820.97	0.17		4,730	769.42	0.162	
9. Total Variation	5,279	1,307.10			2,519	610.47			5,279	1,298.38			5,279	1,295.59		

* $P < .05$

n.s. $P > .05$

Table 15

Summary of Findings with Respect to Group and Candidate Differences on Form B

Source of Variation	Groups Compared			
	B vs. A Texas Females: Black vs. White	G vs. F Illinois Females: Black vs. White	B vs. E Texas Females: Black vs. White	C vs. D Alabama Females: Black vs. White
Interactions, Groups by Item Types:				
(1) Groups by (Black vs. Modern Items)	*	*	*	*
(2) Groups by (Traditional vs. Black and Modern Items)	*	*	*	*
(3) Groups by (Literature vs. Art Items)	*	n.s.	n.s.	n.s.
(4) Groups by (B X S Items)	n.s.	n.s.	*	n.s.
Interactions, Candidates in Same Group by Item Types:				
(5) Candidates by (Black vs. Modern Items)	n.s.	n.s.	n.s.	*
(6) Candidates by (Traditional vs. Black and Modern Items)	n.s.	*	n.s.	n.s.
(7) Candidates by (Literature vs. Art Items)	n.s.	n.s.	n.s.	*
(8) Candidates by (Social Studies vs. Art and Literature Items)	n.s.	*	*	*
(9) Candidates by (T X L Items)	*	n.s.	*	n.s.
(10) Candidates by (T X S Items)	n.s.	n.s.	*	*

* $P < .05$ (significant, .05 level)

Table 16

Significant Group Differences Related to Cultural Reference of Items[3] on Form B

Composition of Test	Differences Between Group Mean Numbers Right			
	B vs. A Texas Females: Black vs. White	G vs. F Illinois Females: Black vs. White	B vs. E Texas Females: Black vs. White	C vs. D Alabama Females: Black vs. White
65 Black Items	+0.3	+10.5	+3.2	+3.2
65 Modern Items	-19.9	-8.8	-20.2	-19.8
Difference, B-M	+20.2*	+19.3*	+23.4*	+23.0*
65 Items, Half Black, Half Modern	-9.8	+0.8	-8.5	-8.3
65 Traditional Items	-20.4	-7.6	-17.4	-18.2
Difference, (B+M)-T	+10.6*	+8.4*	+8.9*	+9.9*
Difference, B-T (Not tested)	+20.7	+18.1	+20.6	+21.4

* $P < .05$
n.s. $P > .05$